



On a software reliability model by Koch and Spreij

James Ledoux

► To cite this version:

James Ledoux. On a software reliability model by Koch and Spreij. *Statistics and Probability Letters*, 2008, 78 (7), pp.904-914. 10.1016/j.spl.2007.10.002 . hal-00359801

HAL Id: hal-00359801

<https://hal.science/hal-00359801>

Submitted on 26 Aug 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On a software reliability model by Koch and Spreij

James Ledoux*

11 October 2007

Abstract

In a very concise paper, Koch and Spreij have proposed a software reliability model which generalized Jelinski-Moranda's one. The basic assumption is that the failure intensity is proportional to the residual number of failures, but the initial number of failures is not known. In fact, this paper contained original ideas which were rediscovered by others some time later usually in a Bayesian modelling framework. Their paper will be revisited here. Specifically the results supporting the discussion on various specific models by Koch and Spreij (1983) are stated and proved.

KEYWORD: Filtering, Bayesian modelling, Reliability

1 Introduction

Koch and Spreij (1983) have proposed general software reliability models. Let us introduce their basic model in which the assumptions were as follows:

1. each detected fault is immediately corrected;
2. each correction introduces no new fault;
3. the failure intensity is proportional to the residual number of faults.

These assumptions are common to a large amount of software reliability models (Pham, 2000; Singpurwalla and Wilson, 1999). Let us consider a complete filtered probability space $(\Omega, \mathbb{P}, \mathbb{F})$ where $\mathbb{F} := \{\mathbb{F}_t\}_{t \geq 0}$ is a continuous to the right filtration or history such that the counting process $\{N_t\}_{t \geq 0}$ is adapted. This counting process has the Doob-Meyer representation

$$N_t = \int_0^t \lambda_s ds + M_t, \quad N_0 = 0 \quad (1.1)$$

where $\{M_t\}_{t \geq 0}$ is assumed to be an \mathbb{F} -martingale. This is valid if $\mathbb{E}[N_t] < +\infty$ for every t . From Assumption 3, the random process $\{\lambda_t\}_{t \geq 0}$ is a stochastic intensity of $\{N_t\}_{t \geq 0}$ which has the form

$$\lambda_t = \nu R_t = \nu (R_0 - N_t) \quad (1.2)$$

*IRMAR UMR-CNRS 6625 & Institut National des Sciences Appliquées, 20 avenue des Buttes de Coësmes, 35043 Rennes Cedex FRANCE, james.ledoux@insa-rennes.fr

where R_t is the number of faults still contained in the software at time t and ν is a (random) factor of proportionality. Note that R_0 is the initial number of faults contained in the software. ν may be interpreted as a fault-manifestation rate.

Koch and Spreij have discussed how martingale theory provides an unique framework to analyze software reliability models through the basic model above and some extensions, one of them being discussed in Section 3. Their paper has been largely ignored in the literature on software reliability though many basic ideas in later developments on Generalized Order Statistics (GOS) models, on connections between GOS and Non-Homogeneous Poisson Process (NHPP) models, on Bayesian inference for software reliability models were already included (e.g. Langberg and Singpurwalla, 1985; Miller, 1986; Kuo and Yang, 1996; Wang, 2005). The point was to determine the joint distribution of the number of residual faults and the manifestation fault rate conditionally to the history of the failure process. This was done in solving a filtering problem. In other words, they obtained the posterior distribution of the parameters of the model as a function of their initial distributions. Then, a large number of now well established models can be obtained using various initial distributions. Why this paper has been ignored? Maybe, one of the main reason is that it is a very concise paper which only gives a flavour of results. Specifically, the reader is reported to (Koch and Spreij, 1982) for proofs and most of technical results. The aims of this paper are: first, give a derivation of the main result; second, point out the main interests of the present framework for software reliability modelling.

2 One class of faults

It follows from Assumptions 1 and 2 that the jump process $\{R_t\}_{t \geq 0}$ has the semi-martingale decomposition

$$R_t = - \int_0^t \lambda_s ds - M_t \quad (2.1)$$

and R_0 has some probability distribution. Thus, the dynamics of the couple of jump processes $\{N_t\}_{t \geq 0}$ and $\{R_t\}_{t \geq 0}$ are given by (1.1) and (2.1) respectively. Now, the Doob-Meyer decomposition of $\{N_t\}_{t \geq 0}$ with respect to its internal history $\{\mathbb{F}_t^N\}_{t \geq 0}$ is for $t > 0$

$$N_t = \int_0^t \widehat{\lambda}_s ds + \widehat{M}_t \quad \text{where } \widehat{\lambda}_t = \mathbb{E}[\lambda_t | \mathbb{F}_t^N] = \mathbb{E}[\nu R_t | \mathbb{F}_t^N]. \quad (2.2)$$

In this section, the conditional expectation $\mathbb{E}[Z_t | \mathbb{F}_t^N]$ is denoted by \widehat{Z}_t for any integrable \mathbb{F} -adapted random process $\{Z_t\}_{t \geq 0}$. Moreover, $\{M_t\}_{t \geq 0}$, $\{\widehat{M}_t\}_{t \geq 0}$ will denote a generic \mathbb{F} -martingale and \mathbb{F}^N -martingale respectively.

2.1 Posterior distribution

Let us assume that the probability distribution of (ν, R_0) has a density f_0 with respect to the measure μ on $\mathbb{R}_+ \times \mathbb{N}$, product of the Lebesgue measure on \mathbb{R}_+ and the counting measure on \mathbb{N} . The next lemma states that the probability distribution of (ν, R_t) conditional to \mathbb{F}_t^N has a density f_t with respect to μ that is a solution of a stochastic differential equation. In connection with the fact that \mathbb{F}_t^N corresponds to the observed data, this conditional distribution is called the *posterior distribution* of (ν, R_t) and f_t the corresponding posterior density.

Lemma 1 *The quantity $\widehat{\lambda}_s^+$ denotes $1/\widehat{\lambda}_s$ if $\widehat{\lambda}_s > 0$ and 0 otherwise. Then the posterior density of (ν, R_t) for $t > 0$ satisfies*

$$\begin{aligned} f_t(\nu, r) = & f_0(\nu, r) + \int_0^t [(r+1)\nu f_s(\nu, r+1) - r\nu f_s(\nu, r)] ds \\ & + \int_0^t \widehat{\lambda}_{s-}^+ \left[(r+1)\nu f_{s-}(\nu, r+1) - \widehat{\lambda}_{s-} f_{s-}(\nu, r) \right] (dN_s - \widehat{\lambda}_{s-} ds) \end{aligned} \quad (2.3)$$

Proof. First, note that

$$1_{R_t=r} = 1_{R_0=r} + \sum_{0 < s \leq t} (1_{R_s=r} - 1_{R_{s-}=r}). \quad (2.4)$$

Next, we have with $\Delta R_s := R_s - R_{s-}$ and $\Delta N_s := N_s - N_{s-}$

$$1_{R_s=r} - 1_{R_{s-}=r} = -(1_{R_s=r} - 1_{R_{s-}=r})\Delta R_s = (1_{R_s=r} - 1_{R_{s-}=r})\Delta N_s$$

from Assumption 1. Then

$$\begin{aligned} \sum_{0 < s \leq t} (1_{R_s=r} - 1_{R_{s-}=r}) &= \int_0^t (1_{R_s=r} - 1_{R_{s-}=r}) dN_s = \int_0^t (1_{R_{s-}=r+1} - 1_{R_{s-}=r}) dN_s \\ &= \int_0^t (1_{R_{s-}=r+1} - 1_{R_{s-}=r}) \lambda_s ds \\ &\quad + \int_0^t (1_{R_{s-}=r+1} - 1_{R_{s-}=r}) (dN_s - \lambda_s ds). \end{aligned}$$

The last integral in the right hand side of the equation above defines an \mathbb{F} -martingale since $(1_{R_{s-}=r+1} - 1_{R_{s-}=r})$ is \mathbb{F} -predictable and bounded. Thus, we obtain an \mathbb{F} -semimartingale representation of the process $\{1_{R_t=r}\}_{t \geq 0}$

$$1_{R_t=r} = 1_{R_0=r} + \int_0^t (1_{R_{s-}=r+1} - 1_{R_{s-}=r}) \lambda_s ds + M_t.$$

Since $\lambda_s = \nu R_s$, the final form is

$$1_{R_t=r} = 1_{R_0=r} + \int_0^t ((r+1)1_{R_{s-}=r+1} - r1_{R_{s-}=r}) \nu ds + M_t.$$

Multiplying the previous equality by $1_{\nu \in B}$, for any Borel set B , we obtain

$$1_{\nu \in B, R_t=r} = 1_{\nu \in B, R_0=r} + \int_0^t ((r+1)1_{\nu \in B, R_{s-}=r+1} - r1_{\nu \in B, R_{s-}=r}) \nu ds + 1_{\nu \in B} M_t.$$

Since $1_{\nu \in B}$ is \mathbb{F} -measurable, $\{1_{\nu \in B} M_t\}_{t \geq 0}$ is still an \mathbb{F} -martingale and we have the following \mathbb{F} -semimartingale representation for $\{p_t(B, r) := 1_{\nu \in B, R_t=r}\}_{t \geq 0}$

$$p_t(B, r) = p_0(B, r) + \int_0^t ((r+1)p_{s-}(B, r+1) - rp_{s-}(B, r)) \nu ds + M_t. \quad (2.5)$$

Recall that for any integrable process $\{Z_t\}_{t \geq 0}$ such that $\int_0^t \mathbb{E}[|Z_s|] ds < +\infty$, we have that

$$\widehat{\int_0^t Z_s ds} - \int_0^t \widehat{Z_s} ds$$

defines an \mathbb{F}^N -martingale. Then, taking the conditional expectation of (2.5), we obtain

$$\widehat{p}_t(B, r) = \widehat{p}_0(B, r) + \int_0^t \{(r+1)\widehat{\nu p}_{s-}(B, r+1) - r\widehat{\nu p}_{s-}(B, r)\} ds + \widehat{M}_t. \quad (2.6)$$

Note that $\Delta \widehat{p}_s(B, r) := \widehat{p}_s(B, r) - \widehat{p}_{s-}(B, r) = \Delta \widehat{M}_s$. We know that the \mathbb{F}^N -martingale $\{\widehat{M}_t\}_{t \geq 0}$ has the following representation (Bremaud, 1981)

$$\widehat{M}_t = \int_0^t G_s(B, r)(dN_s - \widehat{\nu R}_{s-} ds) \quad (2.7)$$

where $\{G_t\}_{t \geq 0}$ is the so-called innovation gain. In order to obtain an explicit form of (2.6), it remains to determine the innovation gain. The basic idea is to derive two \mathbb{F}^N -semimartingale representations of the process $\{N_t \widehat{p}_t(B, r)\}_{t \geq 0}$ and then to identify the locally finite variations part of these decompositions since $\{N_t \widehat{p}_t(B, r)\}_{t \geq 0}$ is a special semimartingale.

We derive a first \mathbb{F}^N -semimartingale representation of $\{N_t \widehat{p}_t(B, r)\}_{t \geq 0}$. An integration by parts and using (2.2) and (2.7), we obtain:

$$\begin{aligned} N_t \widehat{p}_t(B, r) &= \int_0^t N_{s-} d\widehat{p}_s(B, r) + \int_0^t \widehat{p}_{s-}(B, r) dN_s + \sum_{s \leq t} \Delta N_s \Delta \widehat{p}_s(B, r) \\ &= \int_0^t N_{s-} d\widehat{p}_s(B, r) + \int_0^t \widehat{p}_{s-}(B, r) \widehat{\lambda}_{s-} ds + \widehat{M}_t + \int_0^t G_s(B, r) dN_s \\ &= \int_0^t N_{s-} d\widehat{p}_s(B, r) + \int_0^t \{\widehat{p}_{s-}(B, r) + G_s(B, r)\} \widehat{\lambda}_{s-} ds + \widehat{M}_t. \end{aligned}$$

We deduce the following final expression from (2.6)

$$\begin{aligned} N_t \widehat{p}_t(B, r) &= \int_0^t N_{s-} \{(r+1)\widehat{\nu p}_{s-}(B, r+1) - r\widehat{\nu p}_{s-}(B, r)\} ds \\ &\quad + \int_0^t \{\widehat{p}_{s-}(B, r) + G_s(B, r)\} \widehat{\lambda}_{s-} ds + \widehat{M}_t. \end{aligned} \quad (2.8)$$

A second representation of $\{N_t \widehat{p}_t(B, r)\}_{t \geq 0}$ results from a “projection” on \mathbb{F}_t^N of the following \mathbb{F} -representation of $\{N_t p_t(B, r)\}_{t \geq 0}$. An integration by parts gives from (1.1) and (2.5)

$$\begin{aligned}
N_t p_t(B, r) &= \int_0^t N_{s-} dp_s(B, r) + \int_0^t p_{s-}(B, r) dN_s + \sum_{s \leq t} \Delta N_s \Delta p_s(B, r) \\
&= \int_0^t N_{s-} ((r+1)p_{s-}(B, r+1) - r p_{s-}(B, r)) \nu ds + M_t \\
&\quad + \int_0^t p_{s-}(B, r) \nu R_{s-} ds + M_t \\
&\quad + \int_0^t (p_{s-}(B, r+1) - p_{s-}(B, r)) \nu R_{s-} ds + M_t \\
&= \int_0^t N_{s-} ((r+1)p_{s-}(B, r+1) - r p_{s-}(B, r)) \nu ds \\
&\quad + \int_0^t p_{s-}(B, r+1) \nu R_{s-} ds + M_t.
\end{aligned}$$

Next, take the conditional expectation with respect to \mathbb{F}_t^N of both sides

$$\begin{aligned}
N_t \widehat{p}_t(B, r) &= \int_0^t N_{s-} ((r+1) \widehat{\nu p}_{s-}(B, r+1) - r \widehat{\nu p}_{s-}(B, r)) ds \\
&\quad + \int_0^t (r+1) \widehat{\nu p}_{s-}(B, r+1) ds + \widehat{M}_t.
\end{aligned} \tag{2.9}$$

We identify the bounded variation parts of formulas (2.8) and (2.9) (that is the Lebesgue integrals):

$$\int_0^t \{\widehat{p}_{s-}(B, r) + G_s(B, r)\} \widehat{\lambda}_{s-} ds = \int_0^t (r+1) \widehat{\nu p}_{s-}(B, r+1) ds.$$

Therefore, for given $B \in \mathcal{B}(\mathbb{R}_+)$ and $r \in \mathbb{N}$, we obtain that the gain is defined by

$$G_s(B, r) = \widehat{\lambda}_{s-}^+ \left[(r+1) \widehat{\nu p}_{s-}(B, r+1) - \widehat{\lambda}_{s-} \widehat{p}_{s-}(B, r) \right]. \tag{2.10}$$

It results from (2.6) with the expression of the gain above, that the density f_t satisfies equation (2.3). \blacksquare

Let us denote the observed failure times by $t_1 < t_2 < \dots$. Equation (2.3) reads as follows: between two jumps of the counting process $\{N_t\}_{t \geq 0}$ the posterior density satisfies the ordinary differential equation on $]t_{n-1}, t_n[$

$$\frac{d}{dt} f_t(v, r) = [\widehat{\lambda}_t - rv] f_t(v, r) \tag{2.11a}$$

with initial condition $f_{t_{n-1}}(v, r)$ and at jump instants $T_n = t_n$

$$\begin{aligned}
f_{t_n}(v, r) - f_{t_n-}(v, r) &= \widehat{\lambda}_{t_n-}^+ \left[(r+1) v f_{t_n-}(v, r+1) - \widehat{\lambda}_{t_n-} f_{t_n-}(v, r) \right] \\
\text{or } f_{t_n}(v, r) &= \widehat{\lambda}_{t_n-}^+ (r+1) v f_{t_n-}(v, r+1).
\end{aligned} \tag{2.11b}$$

where $\widehat{\lambda}_t = \mathbb{E}[\nu R_t \mid \mathbb{F}_t^N] = \int v r f_t(v, r) d\mu$. Then, we can show the following theorem stated by Koch and Spreij (1983).

Theorem 2 *The notation are as in Lemma 1 and the observed failure times are denoted by $t_1 < t_2 < \dots$. The posterior distribution of (ν, R_t) has the density*

$$f_t(v, r) = c(N_t, t) \times f_0(v, N_t + r) \frac{(r + N_t)!}{r!} v^{N_t} \exp \left[-v \left(rt + \sum_{i=1}^{N_t} t_i \right) \right] \quad (2.12a)$$

where $c(N_t, t)$ is given by

$$1/c(N_t, t) = \mathbb{E}_{f_0} \left[\nu^{N_t} \frac{R_0!}{(R_0 - N_t)!} \exp \left[-\nu \left((R_0 - N_t)t + \sum_{i=1}^{N_t} t_i \right) \right] 1_{R_0 \geq N_t} \right]. \quad (2.12b)$$

For any $t > 0$, the quantity $c(N_t, t)$ is just a normalizing term for the right hand side member of the equation (2.12a) to be a probability density on $\mathbb{R}_+ \times \mathbb{N}$ with respect to μ .

Proof. Let us check that the function defined in (2.12a) satisfies equation (2.11a) between two jumps of the counting process. First, we note from (2.12a) that

$$\begin{aligned} \widehat{\lambda}_t &= \int v r f_t(v, r) d\mu \\ &= \int v r c(N_t, t) \times f_0(v, N_t + r) \frac{(r + N_t)!}{r!} v^{N_t} \exp \left[-v \left(rt + \sum_{i=1}^{N_t} t_i \right) \right] d\mu \\ &= c(N_t, t) \\ &\quad \times \mathbb{E}_{f_0} \left[(R_0 - N_t) \nu^{N_t+1} \frac{R_0!}{(R_0 - N_t)!} \exp \left[-\nu \left((R_0 - N_t)t + \sum_{i=1}^{N_t} t_i \right) \right] 1_{R_0 \geq N_t} \right] \\ &= \frac{c(N_t, t)}{c(N_t + 1, t)}. \end{aligned} \quad (2.13)$$

Second, using Lebesgue's theorem, we obtain from (2.12b)

$$\begin{aligned} \frac{dc(N_t, t)}{dt} &= c(N_t, t)^2 \\ &\quad \times \mathbb{E}_{f_0} \left[(R_0 - N_t) \nu^{N_t+1} \frac{R_0!}{(R_0 - N_t)!} \exp \left[-\nu \left((R_0 - N_t)t + \sum_{i=1}^{N_t} t_i \right) \right] 1_{R_0 \geq N_t} \right] \\ &= c(N_t, t) \widehat{\lambda}_t. \end{aligned}$$

Now, it is easily checked that (2.12a) is solution to (2.11a) between two jumps of $\{N_t\}_{t \geq 0}$.

It remains to verify that f_t defined in (2.12a) also satisfies equation (2.11b) at jump instants. A direct inspection shows from (2.12a) and (2.13) that:

$$\begin{aligned} f_{t_n}(v, r) &= c(n, t_n) f_0(v, n+r) \frac{(r+n)!}{r!} v^n \exp \left[-v(rt + \sum_{i=1}^n t_i) \right] \\ f_{t_n-}(v, r+1) &= c(n-1, t_n-) f_0(v, n+r) \frac{(r+n)!}{(r+1)!} v^{n-1} \exp \left[-v(rt_n + \sum_{i=1}^n t_i) \right] \\ \widehat{\lambda}_{t_n-} &= \frac{c(n-1, t_n-)}{c(n, t_n-)} = \frac{c(n-1, t_n-)}{c(n, t_n)}. \end{aligned}$$

The proof is easily completed. ■

2.2 Some specific distributions of (ν, R_0)

If ν and R_0 are deterministic, i.e. f_0 is a Dirac distribution concentrated on a couple of constants (v_0, N) , the posterior distribution is also a Dirac distribution on $(v_0, N - N_t)$ from (2.12a). In such a case, the intensity with respect to \mathbb{F}_t^N is obviously $\widehat{\lambda}_t = v_0(N - N_t)$ and we obtain the celebrated model by Jelinski and Moranda (1972). Now, let us assume that ν is deterministic but R_0 has a Poisson distribution with parameter m , i.e.

$$f_0(v, r) = \delta_{v_0}(v) \exp(-m) \frac{m^r}{r!}.$$

We obtain by direct inspection

$$\frac{1}{c(N_t, t)} = (mv_0)^{N_t} \exp \left[-m(1 - \exp(-v_0 t)) - v_0 \sum_{j=1}^{N_t} t_j \right].$$

We deduce from (2.12a) that R_t has posterior distribution $\text{Pois}(m \exp[-v_0 t])$. Therefore, $\widehat{\lambda}_t = v_0 \mathbb{E}[R_t \mid \mathbb{F}_t^N] = v_0 m \exp(-v_0 t)$ is deterministic and $\{N_t\}_{t \geq 0}$ a Poisson process. We retrieve a model by Goel and Okumoto (1979). This fact was rediscovered by Langberg and Singpurwalla (1985).

3 A finite number of classes of faults

In this section, we consider a second model by Koch and Spreij (1983) in which a family of $K < \infty$ different classes of faults are contained in a software. Each class has the properties of Section 1 with corresponding parameters $N_t^{(j)}, R_t^{(j)}$ and $\nu^{(j)}$. In particular, $\{N_t^{(j)}\}_{t \geq 0}$ has the decomposition

$$N_t^{(j)} = \int_0^t \lambda_s^{(j)} ds + M_t^{(j)} \quad N_0^{(j)} = 0 \quad (3.1)$$

where $\lambda_s^{(j)} := \nu R_s^{(j)}$. In addition, the martingales $\{M_t^{(j)}\}_{t \geq 0}$ are assumed to be independent. Let us introduce the multivariate random variables $\mathbf{N}_t := (N_t^{(j)})_{j=1}^K$, $\mathbf{M}_t := (M_t^{(j)})_{j=1}^K$ and $\mathbf{R}_t := (R_t^{(j)})_{j=1}^K$. As in Section 2, the probability distribution of $(\boldsymbol{\nu}, \mathbf{R}_0)$ is assumed to have a

density f_0 with respect to the measure $\boldsymbol{\mu}$ on $\mathbb{R}_+^K \times \mathbb{N}^K$, product of the Lebesgue measure on \mathbb{R}_+^K and the counting measure on \mathbb{N}^K . Let \mathbf{e}_j be the j th vector of the canonical basis of \mathbb{R}^K . The results of this section are not stated by Koch and Spreij (1983) (except Theorem 8). Koch and Spreij (1982) quoted the existence of Theorem 4.

3.1 Posterior distributions

3.1.1 The multivariate counting process is observed

In a first step, the multivariate $\mathbf{N}_t := (N_t^{(j)})_{j=1}^K$ is considered to be observed, and we are interested in the posterior distribution of $(\boldsymbol{\nu}, \mathbf{R}_t)$ with respect to the internal history $\mathbb{F}_t^{\mathbf{N}}$ of $\{\mathbf{N}_t\}_{t \geq 0}$. The next lemma is a multivariate version of Lemma 1 using the decompositions (3.1) in place of (1.1). The proof is omitted.

Lemma 3 *The notation are as in Lemma 1. The posterior density of $(\boldsymbol{\nu}, \mathbf{R}_t)$ for $t > 0$ is a solution of the following equation*

$$\begin{aligned} f_t(\mathbf{v}, \mathbf{r}) &= f_0(\mathbf{v}, \mathbf{r}) + \sum_{j=1}^K \int_0^t [(r_j + 1)v_j f_s(\mathbf{v}, \mathbf{r} + \mathbf{e}_j) - r_j v_j f_s(\mathbf{v}, \mathbf{r})] ds \\ &\quad + \sum_{j=1}^K \int_0^t (\widehat{\lambda}_{s-}^{(j)})^+ \left[(r_j + 1)v_j f_{s-}(\mathbf{v}, \mathbf{r} + \mathbf{e}_j) - \widehat{\lambda}_{s-}^{(j)} f_{s-}(\mathbf{v}, \mathbf{r}) \right] (dN_s^{(j)} - \widehat{\lambda}_s^{(j)} ds) \end{aligned} \quad (3.2)$$

where $\widehat{\lambda}_t^{(j)} = \mathbb{E}[\lambda_t^{(j)} \mid \mathbb{F}_t^{\mathbf{N}}]$.

Now, the unique solution of (3.2) is given in the following theorem.

Theorem 4 *The notation are as in Theorem 2. The posterior distribution of $(\boldsymbol{\nu}, \mathbf{R}_t)$ has the density*

$$\begin{aligned} f_t(\mathbf{v}, \mathbf{r}) &= c(\mathbf{N}_t, t) \times f_0\left(\mathbf{v}, \mathbf{r} + \sum_{j=1}^{N_t} \mathbf{e}_{l_j}\right) \\ &\quad \frac{F(\mathbf{r} + \sum_{j=1}^{N_t} \mathbf{e}_{l_j})}{F(\mathbf{r})} \left[\prod_{j=1}^{N_t} v_{l_j} \right] \exp\left(-t \sum_{j=1}^K v_j r_j + \sum_{j=1}^{N_t} t_j v_{l_j}\right) \end{aligned} \quad (3.3)$$

where

- the l_j 's are such that $N_{t_j}^{(l_j)} - N_{t_j-}^{(l_j)} = 1$;
- $F(\mathbf{r}) = r_1! \cdots r_K!$ for any $\mathbf{r} = (r_1, \dots, r_K) \in \mathbb{N}^K$;
- $c(\mathbf{N}_t, t)$ is a normalizing term for the right hand side member of equation (3.3) to be a probability density on $\mathbb{R}_+^K \times \mathbb{N}^K$ at any time $t > 0$.

3.1.2 Only the aggregated counting process is observed

In a second step, the aggregated number of observed failures

$$N_t := \sum_{j=1}^K N_t^{(j)} \quad (3.4)$$

is assumed to be observed. Doob-Meyer's decompositions with respect to \mathbb{F} and \mathbb{F}^N are from (3.1)

$$N_t = \int_0^t \lambda_s ds + m_t \quad \text{where } \lambda_s := \sum_{j=1}^K \lambda_s^{(j)} = \sum_{j=1}^K \nu^{(j)} R_s^{(j)} \quad (3.5a)$$

and

$$N_t = \int_0^t \widehat{\lambda}_s ds + \widehat{m}_t \quad \text{where } \widehat{\lambda}_s := \sum_{j=1}^K \widehat{\lambda}_s^{(j)} = \sum_{j=1}^K \mathbb{E}[\nu^{(j)} R_s^{(j)} \mid \mathbb{F}_t^N] \quad (3.5b)$$

where $\{m_t\}_{t \geq 0}$, $\{\widehat{m}_t\}_{t \geq 0}$ are an \mathbb{F} -martingale and an \mathbb{F}^N -martingale respectively. Then, the distribution of $(\boldsymbol{\nu}, \mathbf{R}_t)$ conditionally to \mathbb{F}_t^N may be derived as for Lemmas 1 and 3.

Lemma 5 *The notation are as in Theorem 4. The density g_t of the conditional distribution of $(\boldsymbol{\nu}, \mathbf{R}_t)$ with respect to \mathbb{F}_t^N is a solution of the following equation*

$$\begin{aligned} g_t(\mathbf{v}, \mathbf{r}) &= g_0(\mathbf{v}, \mathbf{r}) + \sum_{j=1}^K \int_0^t [(r_j + 1)v_j g_s(\mathbf{v}, \mathbf{r} + \mathbf{e}_j) - r_j v_j g_s(\mathbf{v}, \mathbf{r})] ds \\ &\quad + \int_0^t (\widehat{\lambda}_{s-})^+ \sum_{j=1}^K [(r_j + 1)v_j g_{s-}(\mathbf{v}, \mathbf{r} + \mathbf{e}_j) - \widehat{\lambda}_{s-}^{(j)} g_{s-}(\mathbf{v}, \mathbf{r})] (dN_s^{(j)} - \widehat{\lambda}_s ds) \end{aligned} \quad (3.6)$$

where $\widehat{\lambda}_t$ defined in (3.5b).

Theorem 6 *The conditional distribution of $(\boldsymbol{\nu}, \mathbf{R}_t)$ given \mathbb{F}_t^N has the following density with respect to $\boldsymbol{\mu}$*

$$g_t(\mathbf{v}, \mathbf{r}) = c(t) \sum_{l_1, \dots, l_{N_t}} \frac{f_t(\mathbf{v}, \mathbf{r})}{c(l_1, \dots, l_{N_t}, t)} \quad (3.7)$$

where $f_t(\mathbf{v}, \mathbf{r})$ is given by (3.3) and $c(l_1, \dots, l_{N_t}, t)$ denotes its associated normalizing constant. The quantity $c(t)$ is a global normalizing constant.

3.2 A finite number of classes with at most one fault per class

If we assume that each class of fault contains at most one fault, that is $R^{(j)} \in \{0, 1\}$, the l_j 's in (3.3) are all different and $r_j = 0$ for a class of undetected fault. We deduce the following result from Theorem 4.

Corollary 7 *The notation are as in Theorem 4. The posterior distribution of $(\boldsymbol{\nu}, \mathbf{R}_t)$ has the density*

$$f_t(\mathbf{v}, \mathbf{r}) = c(t) \times f_0\left(\mathbf{v}, \mathbf{r} + \sum_{j=1}^{N_t} \mathbf{e}_{l_j}\right) \left[\prod_{j=1}^{N_t} (r_{l_j} + 1) v_{l_j} \right] \exp\left(-t \sum_{j=1}^K v_j r_j + \sum_{j=1}^{N_t} t_{l_j} v_{l_j}\right)$$

where $c(t)$ is a normalizing term for the right hand side member of the equation above to be a probability density on $\mathbb{R}_+^K \times \mathbb{N}^K$ with respect to $\boldsymbol{\mu}$ at any time $t > 0$.

The posterior distributions of $(\boldsymbol{\nu}, \mathbf{R}_t)$ given in Corollary 7 and Theorem 6 allows us to justify the following properties stated by Koch and Spreij (1983).

Theorem 8 *The notation are as in Theorem 6 and Corollary 7. Assume that the $(\nu^{(j)}, R_0^{(j)})$'s are independent under f_0 . Then,*

1. $(\nu^{(j)}, R_t^{(j)})$'s are also independent under f_t with $t > 0$;
2. if the $(\nu^{(j)}, R_0^{(j)})$'s are identically distributed under f_0 , then the $(\nu^{(j)}, R_t^{(j)})$'s associated with classes of undetected faults are also identically distributed under f_t and g_t .

Moreover, N_t is a sufficient statistics for estimating any function of the pairs $(\nu^{(j)}, R_t^{(j)})$ corresponding to a class of undetected faults conditionally to $\mathbb{F}_t^{\mathbf{N}}$. In particular, we have

$$\widehat{\lambda}_t = \widehat{\lambda}_t = (K - N_t) \widehat{\lambda}_t^{(j)} . \quad (3.8)$$

where j is an index corresponding to any class of undetected faults.

Proof. The probabilistic properties on the posterior distributions are easily deduced from the expressions of f_t and a_t in function of f_0 given in Corollary 7 and Theorem 6. The property on exhaustiveness of the aggregated counting variable results from the fact that $(\nu^{(j)}, R_t^{(j)})$'s associated with classes of undetected faults have the same posterior distribution under f_t and g_t . In particular, $\nu^{(j)} R_t^{(j)}$'s have the same conditional expectations so that (3.8) is valid. ■

If K is assumed to be known and under the assumptions of the second statement of Theorem 6, the multivariate counting process $\{\mathbf{N}_t\}_{t \geq 0}$ has a likelihood given by (Bremaud, 1981)

$$L_t^{(K)} = \exp\left(-\int_0^t \sum_{j=1}^K (\widehat{\lambda}_s^{(j)} - 1) ds + \int_0^t \sum_{j=1}^K \log(\widehat{\lambda}_{s-}^{(j)}) dN_s^{(j)}\right).$$

This reads as follows

$$\begin{aligned}
L_t^{(K)} &= \exp \left(- \int_0^t (\hat{\lambda}_t - 1) ds + (K - 1)t + \sum_{j=1}^{N_t} \log \hat{\lambda}_{t_j-}^{(l_j)} \right) \\
&= \exp \left(- \int_0^t (\hat{\lambda}_t - 1) ds + \sum_{j=1}^{N_t} \log ((K - j + 1) \hat{\lambda}_{t_j-}^{(l_j)}) \right. \\
&\quad \left. + (K - 1)t - \sum_{j=1}^{N_t} \log (K - j + 1) \right) \\
&= \exp \left(- \int_0^t (\hat{\lambda}_t - 1) ds + \int_0^t \log \hat{\lambda}_{s-} dN_s \right) \\
&\quad \times \exp \left((K - 1)t - \int_0^t \log (K - N_{s-}) dN_s \right).
\end{aligned}$$

The first exponential above is the likelihood L_t at time t associated with the aggregated process $\{N_t\}_{t \geq 0}$. Therefore, maximize $L_t^{(K)}$ with respect to some deterministic parameters is equivalent to maximize L_t . This agrees with the results of the previous theorem.

Let us consider the case in which each class has a single fault, i.e. $R_0^{(j)} = 1$ for every j . If each $\nu^{(j)}$ is equal to some positive constant ϕ_j , we obtain the Exponential Order Statistics models developed by Miller (1986). If we consider the $\nu^{(j)}$'s as random variables, we obtain an DS/EOS proposed by Miller (1986). Moreover, if the $\nu^{(j)}$ are identically distributed under f_0 , then we obtain the GOS model, that is the counting process $\{N_t\}_{t \geq 0}$ has the \mathbb{F}^N -intensity

$$\hat{\lambda}_t = (K - N_t) \times \phi(t) \quad (3.9)$$

where $\phi(t) = \mathbb{E}[\nu^{(j)} \mid \mathbb{F}_t^N]$ is the hazard rate of the common prior distribution of the time to detection of any of the K faults. For instance, if $\nu^{(j)}$ has a gamma prior distribution $\Gamma(\alpha, \beta)$, then the posterior distributions are

- $\nu^{(j)} \mid \mathbb{F}_t^N \sim \Gamma(\alpha, \beta + t)$ for a class j of undetected faults at time t and

$$\mathbb{E}[\nu^{(j)} \mid \mathbb{F}_t^N] = \frac{\alpha}{\beta + t}$$

- $\lambda_t \mid \mathbb{F}_t^N \sim \Gamma(\alpha(K - N_t), \beta + t)$ and

$$\hat{\lambda}_t = (K - N_t) \times \frac{\alpha}{\beta + t}.$$

We obtain a model introduced by Littlewood (1981). In particular, this justifies the claim in Littlewood (1981) that $\nu^{(j)}$ has distribution $\Gamma(\alpha, \beta + t)$ conditionally to \mathbb{F}_t^N for any undetected fault. More generally, under the assumptions for statement 2 of Theorem 8 to be true, the intensity $\hat{\lambda}_t$ is characterized by the posteriori distribution of $\nu^{(j)}$ corresponding to any undetected fault. In particular, this is the basic fact for the analysis developed by Wang (2005).

Remark 9 *The case of an infinite number K of classes of faults could be also discussed. In particular, if we consider an eventually infinite number K of classes containing only a single fault with independent and identically distributed $\nu_0^{(j)}$'s, and K is assumed to be random with Poisson distribution, we obtain that the conditional distribution of the residual number of faults to \mathbb{F}_t^N is Poisson with time dependent parameter. Then, we can show that GOS models can be transformed into NHPP models as shown by Kuo and Yang (1996).*

4 Martingale approach in software reliability

We briefly discuss the benefits in using the martingale approach in software reliability from Ledoux (2003); Gaudoin and Ledoux (2007). It is clear that the failure process of a software may basically be modelled by the random point process of the times (of detection) of failure $\{T_n\}_{n \geq 0}$, or equivalently by the failure counting process $\{N_t\}_{t \geq 0}$. In such a framework, the standard way to construct a software reliability model consists in choosing a probability distribution for $\{T_n\}_{n \geq 0}$ or $\{N_t\}_{t \geq 0}$. The history \mathbb{F} with respect to which $\{N_t\}_{t \geq 0}$ is adapted, is a central concept of stochastic modelling. Indeed, it contains basic assumptions under which the model is constructed.

In the first case, let us consider the internal history \mathbb{F}^N . Then, the central modelling assumption is that only the past of the failure process affects the future of the failure process. Next, we have the basic decomposition (2.2) where $\{\hat{\lambda}_t\}_{t \geq 0}$ is the \mathbb{F}^N -intensity of the model. It is well known that $\{\hat{\lambda}_t\}_{t \geq 0}$ has an explicit form (Jacod, 1975):

$$\hat{\lambda}_t = \sum_{n \geq 0} h_{X_{n+1} | \mathbb{F}_{T_n}^N} (t - T_n) 1_{\{T_n < t \leq T_{n+1}\}} \quad (4.1)$$

where $\mathbb{F}_{T_n}^N = \sigma(T_n, \dots, T_1)$ and $h_{X_{n+1} | \mathbb{F}_{T_n}^N}(\cdot)$ is the hazard rate function of the inter-failure distribution $X_{n+1} := T_{n+1} - T_n$ conditionally to T_n, \dots, T_1 . It is easily seen from a display of a typical path of $\hat{\lambda}_t$ that process $\{\hat{\lambda}_t\}_{t \geq 0}$ is a concatenation of the hazard rate functions of conditional distributions of inter-failure times. A very large collection of software reliability models have been directly developed using specific forms the sequence of hazard rates. This kind of models is known as models of *self-excited point processes* in engineering (Snyder and Miller, 1991) which have been used by Gaudoin (1990) to unify a large number of existing software reliability models. This last point was rediscovered by Chen and Singpurwalla (1997) and is fully developed in chapter 3 of (Singpurwalla and Wilson, 1999). Note that in general, $\hat{\lambda}_t$ only depends on the past of the point process through t, N_t, T_{N_t} (Ledoux, 2003; Singpurwalla and Wilson, 1999; Pham, 2000).

The benefit of the martingale approach in this first case is to provide a mathematically founded framework to develop software reliability models. In particular, the definition of intensity has not to be based on existence of the limit (Snyder and Miller, 1991)

$$\lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \mathbb{P}\{N_{(t+\Delta t)-} - N_{t-} = 1 \mid \mathbb{F}_{t-}^N\}, \quad (4.2a)$$

as well as the “conditional orderliness” condition to hold:

$$\lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \mathbb{P}\{N_{(t+\Delta t)-} - N_{t-} \geq 1 \mid \mathbb{F}_{t-}^N\} = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \mathbb{P}\{N_{(t+\Delta t)-} - N_{t-} = 1 \mid \mathbb{F}_{t-}^N\} \quad (4.2b)$$

Existence of such limits introduces unnecessary technical subtleties, and conditions, asserting that (4.2a) and (4.2b) hold, must be stated (Chen and Singpurwalla, 1997; Singpurwalla and Wilson, 1999).

The second case concerns what we call “conditionally self-exciting point processes”, where some background information \mathcal{H}_0 is added to the initial state of the model. In a more formal term, the considered history \mathbb{F} is defined by $\mathbb{F}_t := \mathbb{F}_t^N \vee \mathcal{H}_0$ and, in general, \mathcal{H}_0 is some set of parameters involved in the conditional probability distributions. In some sense, this is the basic framework when we deal with bayesian inference for the parameters of classic self-excited models. In fact, formula (4.1) also defines the intensity with respect to \mathbb{F} (Jacod, 1975), using the hazard rate of the conditional distribution of X_{n+1} given $T_n, \dots, T_1, \mathcal{H}_0$. The present paper shows that using stochastic calculus (and specifically filtering), we can get mathematically founded formulas for quantities of interest like “ $\mathbb{P}\{\mathcal{H}_0 \mid \sigma(N_t, T_{N_t}, \dots, T_1)\}$ ” which represents of the probability distribution of the random elements in \mathcal{H}_0 conditionally to the past N_t, T_{N_t}, \dots, T_1 of the failure process. Let us mention that such a conditional distribution allows to derive the intensity with respect to \mathbb{F}^N from (4.1) and from formulas of the type:

$$\begin{aligned} f_{X_{n+1}|T_n, \dots, T_1}(x) &= \int f_{X_{n+1}|T_n, \dots, T_1, \mathcal{H}_0}(x) d\mathbb{P}\{\mathcal{H}_0 \mid T_n, \dots, T_1\} \\ \mathbb{P}\{X_{n+1} \geq t \mid T_n, \dots, T_1\} \\ &= \int \mathbb{P}\{X_{n+1} \geq t \mid T_n, \dots, T_1, \mathcal{H}_0\} d\mathbb{P}\{\mathcal{H}_0 \mid T_n, \dots, T_1\} \end{aligned}$$

(see Singpurwalla, 1989, for related discussions).

Finally, there are situations that do not fit to the (conditional) self-exciting modelling. Let us mention two following basic instances:

1. failure processes affected by random environmental factors : think about regression models where covariables are incorporated in the specification of the intensity (e.g. see Gandy and Jensen, 2004; Singpurwalla and Wilson, 1999, Chap. 7);
2. failure processes driven by an auxiliary stochastic model : Özekici and Soyer (2003) consider a model of the operational profile of the software as auxiliary process (see also Singpurwalla et al., 2003); in architecture-based software reliability modelling, the auxiliary process is a random model of the architecture (see Ledoux, 2008, and references therein).

Statistical analysis of such models can be based on the martingale approach with special emphasis on stochastic filtering to deal with posterior distributions and inference with observed partially processes (e.g. see Andersen et al., 1993; Gandy and Jensen, 2005; Ledoux, 2008).

5 Conclusion

The models discussed by Koch and Spreij (1983) give an unified view of some standard software reliability models. Thus, these classes of models may be understood as Bayesian version of the Jelinski-Moranda’s one. This fact was rediscovered under various forms in papers published later. In particular, the explicit form of the posterior distribution introduced the way of

Bayesian estimation of reliability via Markov Chain Monte Carlo methods as done by Kuo and Yang (1996). It should be pointed out that new models could be derived from this framework using various distributions of initial numbers of faults and manifestation fault rates. We hope that the present paper will contribute to promote the position of (Koch and Spreij, 1983) in the literature on software reliability modelling.

References

- Andersen, P. K., Borgan, O., Gill, R. D., Keiding, N., 1993. Statistical models on counting processes. Springer, New York.
- Bremaud, P., 1981. Point Processes and Queues. Springer.
- Chen, Y., Singpurwalla, N. D., 1997. Unification of software reliability models by self-exciting point processes. *Adv. in Appl. Probab.* 29, 337–352.
- Gandy, A., Jensen, U., 2004. A non-parametric approach to software reliability. *Appl. Stoch. Models Bus. Ind.* 20, 3–15.
- Gandy, A., Jensen, U., 2005. On goodness of fit tests for Aalen’s additive risk model. *Scand. J. Statist.* 32, 425–445.
- Gaudoin, O., 1990. Statistical tools for software reliability evaluation. Ph.D. thesis, Université Joseph Fourier - Grenoble I, in french.
- Gaudoin, O., Ledoux, J., 2007. Random modeling in software reliability. Hermès, in french.
- Goel, A. L., Okumoto, K., 1979. Time-dependent error-detection rate model for software reliability and other performance measures. *IEEE Trans. Reliab.* R-28, 206–211.
- Jacod, J., 1975. Multivariate point processes : Predictable projection, radon-nikodym derivatives, representation of martingales. *Z. Wahrsch. Verw. Gebiete* 31, 235–253.
- Jelinski, Z., Moranda, P. B., 1972. Software reliability research. In: Freiberger, W. (Ed.), *Statistical Methods for the Evaluation Of Computer System Performance*. Academic Press, pp. 465–484.
- Koch, G., Spreij, P., 1982. A martingale approach to software reliability. Tech. Rep. 384, Twente University of Technology.
- Koch, G., Spreij, P., 1983. Software reliability as an application of martingale & filtering theory. *IEEE Trans. Reliab.* 32, 342–345.
- Kuo, L., Yang, T. Y., 1996. Bayesian computation for nonhomogeneous Poisson processes in software reliability. *J. Amer. Statist. Ass.* 91 (434), 763–773.
- Langberg, N., Singpurwalla, N., 1985. A unification of some software reliability models. *SIAM J. Sci. Statist. Comput.* 6, 781–790.

- Ledoux, J., 2003. Chap. 12: Software reliability modeling. In: Pham, H. (Ed.), Handbook of Reliability Engineering. Springer, pp. 213–234.
- Ledoux, J., 2008. Point processes in software reliability. In: Huber, C., Limnios, N., Besbah, M., Nikulin, M. (Eds.), Mathematical Methods for Survival Analysis, Reliability and Quality of Life. Hermès/ISTE–John Wiley and sons, pp. 221–239.
- Littlewood, B., 1981. Stochastic reliability-growth : a model for fault-removal in computer programs and hardware designs. IEEE Trans. Reliab. 30, 313–320.
- Miller, D. R., 1986. Exponential order statistic models for software reliability growth. IEEE Trans. Software Eng. 12, 12–24.
- Özekici, S., Soyer, R., 2003. Reliability of software with an operational profile. European Journal of Operational Research 149, 459–474.
- Pham, H., 2000. Software Reliability. Springer.
- Singpurwalla, N. D., 1989. A unifying perspective on statistical modeling. SIAM Review 31, 560–564.
- Singpurwalla, N. D., Mazzuchi, T. A., Ozekici, S., Soyer, R., 2003. Stochastic process models for reliability in dynamic environments. In: Rao, C. R., Khattree, R. (Eds.), Handbook of Statistics: Statistics in Industry. Vol. 22. Elsevier Science, pp. 1109–1129.
- Singpurwalla, N. D., Wilson, S. P., 1999. Statistical Methods in Software Engineering. Springer.
- Snyder, D. L., Miller, M. I., 1991. Random Point Processes in Time and Space. Springer.
- Wang, R., 2005. A mixture and self-exciting model for software reliability. Statist. Probab. Lett. 72, 187–194.